PERFORMANCE EVALUATION OF FEATURE SELECTION METHOD FOR MACHINE LEARNING ALGORITHMS TO DETECT WORMHOLE ATTACK IN MANETS

K. Gayathri [1] & Dr. R. Vidyabanu[2]
Research Scholar[1] & Assistant Professor[2]
Department of Computer Science,
L.R.G. Government Arts College for Women, Tirupur.

Abstract: *A Mobile Ad Hoc Network (MANET) consists of a collection of wireless mobile nodes that forms a temporary network without having any fixed infrastructure or centralized administration. MANET is infrastructure-less, lack of centralized monitoring and dynamic changing network topology. It is highly vulnerable to different attack due to open error prone shared wireless medium, which may lead to severe damages on infrastructures. Hence the intelligent attack detection system needs to be obtaining the lime light of research based on recent machine learning algorithms. This paper introduced the Whale optimized features for machine algorithms to detect the wormhole attacks among the MANET systems. Since wormhole attacks can happen mostly at network layer of OSI Model. The algorithm has been tested on the OMNET++ environment integrated with python tool. Meanwhile the proposed algorithm has been compared with the other existing algorithm such as Particle Swarm Optimization (PSO), Logistic Regression (LR), Random Forest (RF) Classifier, Navie Bayes (NB) and K Nearest Neighbors (KNN), in which the proposed idea has outperformed the other intelligent algorithms in terms of accuracy, precision, recall and f1-score.*

Keywords: MANET, Wormhole, OMNET, Feature selection, Machine Learning

1. Introduction

A mobile Ad hoc network (MANET) is a collection of two or more devices or nodes equipped with wireless communication and networking capabilities. These node includes laptop, computers, wireless phones and so on, have a limited transmission range. Such a wireless ad-hoc network is infrastructure less, self-organizing, adaptive and does not require any centralized administration. If two such devices are located within transmission range of each other, they can communicate directly. Each node can communicate directly with only few nodes within the communication range and has to forward messages using the neighbor nodes until the messages arrive at the destination nodes. Meanwhile the transmission among sender and receiver can utilize numerous nodes as intermediate nodes, many routing protocols have been proposed for the MANETS. Most of the protocol assumes that other nodes are trustable so they do not consider the security and attack issues. The lack of infrastructure, rapid deployment practices, and the hostile environments in which MANETS are deployed make them vulnerable to a wide range of security attacks. However most of these attacks are performed by a single malicious node. Many solutions exist to solve single node attacks but they cannot prevent from the attacks that are executed by colluding malicious node such as wormhole attack [1].

Wormhole attack is more dangerous than single node attacks. In a wormhole attack, an attacker connects two distant points in the network, and then replays them into the network from that point. An example is shown in Fig. 1. Here S and D are the two end-points of the wormhole link. In this figure, wormhole attack is that all the nodes in area ‘A’ assume that nodes in area ‘B’ are their neighbors and vice versa.

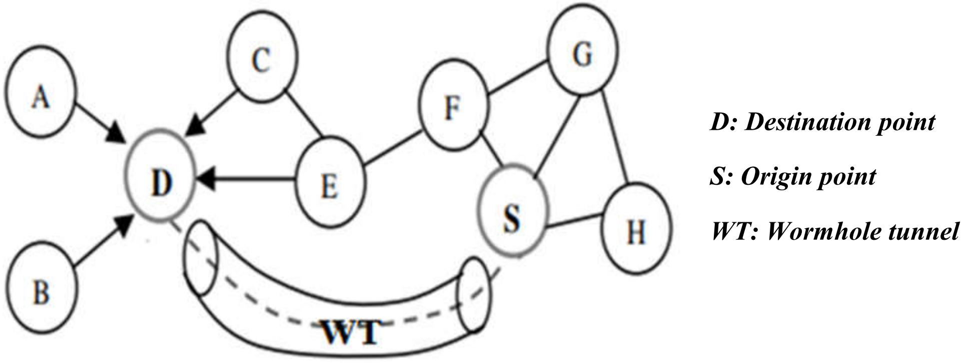


Fig1. Wormhole attack in a network

Machine learning and Deep learning are playing a very important role in the classification and prediction of the data. To design, an intelligent intrusion system we need the most powerful algorithm to detect and predict the attacks in accordance to take its countermeasures. . Several techniques such as machine learning algorithms were used for designing an intelligent intrusion system, but detection and measurement need improvisation. [2].

The paper focuses on the implementation of most powerful machine learning algorithms integrated with the whale optimization to detect the wormhole attacks in MANET. The proposed system consists of three different stages such as Data collection, Feature Extraction, Feature Selection and Prediction of attacks. The MANET network is simulated by OMNET Environment and tested with the proposed algorithms in order to measure the accuracy of detection. Moreover the paper also details about the comparative analysis between the proposed and other existing machine learning algorithms in detection of different attacks.

The organization of the paper is as follows: Section-II deals with overview of Literature Review. Section -III illustrates with the System Model. The data collection unit is detailed in Section-IV. Section -V deals with the experimental setup, performance evaluation and comparative analysis between the different learning algorithms. Finally conclusion is presented in Section-VI.

2. Literature Review

In literature, many researchers were concerned with the security of the MANET. They proposed many machine learning algorithms to detect security threats and find a model to prevent their consequences. Some existing works are described in this section.

Holden [3] has proposed hybrid PSO algorithm that can deal with nominal attributes without going for the both conversion and nominal attribute values. To overcome the drawback (features) that the PSO/ACO algorithm lacks, the proposed method shows simple rule set efficiently to increase in accuracy.

Shaon & Ferens [4] proposed a computationally intelligent approach to detect the wormhole attacks. In this approach, an Artificial Neural Network (ANN) was proposed for detecting the possible locations of wormhole nodes in both equally and non-equally distributed sensor networks. Nonetheless, an exact location of wormhole nodes was not efficiently detected.

Grover et al., [5] proposed ML-based approaches to classify the nodes behavior, i.e. whether the communicating nodes in a vehicular plane is honest or malicious. They implemented different types of misbehaviors and used Naïve Bayes, IBK, J-48, Random Forest and Ada Boost classifiers to classify the behavior.

Ardjani [6] applied SVM with PSO as (PSO-SVM) to optimize the performance of SVM. 10-fold cross-validation is done to estimate the accuracy. It utilizes the advantage of minimum structural risk with global optimizing features. The result shows better accuracy with high execution time.

3. System Model

In this section, we discuss preliminaries of our research work that includes simulation tool (OMNET++), wormhole attack, Feature Selection Algorithms and ML Algorithms.

3.1 OMNET++

OMNeT++ is a discreet event simulator based on modular and object-oriented architecture. OMNeT++ is not a concrete simulator, but provides infrastructure and tools to create simulations. OMNeT++ has a component based architecture and each component created is reusable for other simulations. Internal Architecture of OMNeT++ consists of:

1) SIM is simulation kernel and class library. Sim is a library that connected with program simulation.

2) Envir is a library that consists of general code for all interfaces. Main procedure main() exists to provide services like INI file handling or specific implementation for specific interfaces. Envir shows in Sim and execution model trough facade object named EV.

3) Cmdenv & Tkenv is a specific interface implementation. Simulation can be connected with Cmdenv and/or Tkenv.

4) Model Component Library is consists of simple module definition with C++ implementation, compound module, channels, networks, message types and all things that connected with model that contains in simulation.

5) Executing Model is a model created for simulation. This model consists of objects that instantiated from model component library instance.

INET Framework is open-source extension that allows OMNeT++ to simulate network communication in TCP/IP architecture. INET Framework consists of protocol e.g. IPv4, IPv6, TCP, SCTP, UDP, PPP, Ethernet, and IEEE 802.11. For traffic engineering simulation, INET Framework supports various protocols e.g. IntServ, DifServ, MPLS, RSVP-TE and LDP Signaling [7].

3.2 Wormhole Attack

Wormhole attack is also called as tunneling attack. It is one of the sophisticated and severe attacks in MANET. In this attack the conspiring nodes creates the tunnel among the two nodes for transmitting the packets, appealing that it offers the shortest path to the destination and take full control of the node. The Wormhole can drop the packets by short-circuiting the systematic flow of the routing packets or it can be wisely transmit the packets to avoid detection.

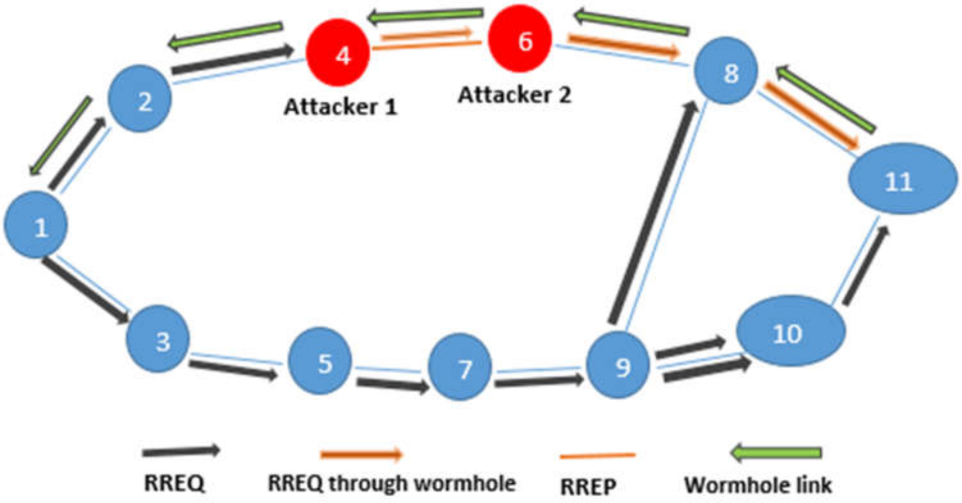


Fig.2. Wormhole attack system

In the above figure, we assume that the 4 and 6 nodes are the colluding attackers and the node 1 i.e. source node is mark to be attacked. During the attack, the source node 1 will transmit the route request (RREQ) to the nearby nodes for discovery a route to the destination node 11, its neighbor 2 and 3 onward RREQ to next door neighbor. However the node 4 will record the forwarded RREQ from node 4 and tunnels the RREQ to its colluding partner 6. Then the node 6 will rebroadcast this RREQ to its neighbor 8.Since this RREQ is passed through high speed channel, the request will reach to the destination 11 node. Therefore node 11 will chose the 11-8-2-1 route to unicast the route reply (RREP) to the source node and disregard the same RREQ that arrived later. As a result source once received the RREP will start transmitting the packet through 4 and 6 nodes [8].

3.3 Feature Selection Algorithms

This section deals with the optimizers for an effective feature selection. To discuss the working mechanism of the proposed optimizer, the preliminary background of PSO and whale algorithms is presented here.

3.3.1 Particle Swarm Optimization (PSO)

PSO is a randomly determined optimization technique copied from flocks of birds or else schools of fish. The flock of birds (swarm) has learnt a co-operative method to discover food and every bird in the swarm, changes the hunt model according to their learning knowledge. The concept of PSO algorithm is related to evolutionary algorithm and swarm artificial life systems.

The particles in the swarm (Birds) openly fly over the multidimensional search space. Through the trip, every particle creates its individual velocity along with location. By updating of each particle the entire population is updated. The swarm arrangement drives itself, to move toward the point of upper target function value and in the end the particles assemble around this point [9].

The steps of particle swarm optimization is as follows:

Step 1: Initialization – The swarm particles lie within the pre-defined ranges of velocity and position.

Step 2: Velocity Updating – At every cycle the speeds of the swarm particles are calculated by equation 3.

$$\vec{V_i} = \mathrm{W}\vec{V_i} + c_1 R_1(\vec{P}_{i,best} - \vec{P_i}) + c_2 R_2(\vec{g}_{i,best} - \vec{P_i})$$

$\vec{V_i}$= velocity of particle 'i'
$\vec{P}_{i,best}$= finest position reached by the particle
$\vec{g}_{i,best}$ =best location remembered by the particle individual
'W' = parameter controlling the flying elements
R_1, R_2 =random numbers among 0 and 1
c_1, c_2 =cognitive learning factor and social learning factor

The inclusion of variables of each particle gives the PSO, the facility of correctness in searching. The weighing aspects c_1, c_2 avoid collision among the particles (individuals). After updating particle I, velocity v and random number r is verified also protected in a range indicated, to evade collision.

Step 3: Updating of position – There is an interval among succeeding iterations and hence the positions of the particles undergo change as in below equation.

$$\vec{P_i} = \vec{P_i} + \vec{V_i}$$

After refreshing, $\vec{P_i}$ must be verified and in the allowable range.

Step 4: Updating of memory – Update $\vec{P}_{i,best}$ and $\vec{g}_{i,best}$ using the formula as in below equations,

$$\vec{P}_{i,best} = \vec{P}_i \; if \; f(\vec{P}_i) > f(\vec{P}_{i,best})$$
$$\vec{g}_{i,best} = \vec{g}_i \; if \; f(\vec{g}_i) > f(\vec{g}_{i,best})$$

Where $(\vec{x})$ is the point function subject to extension.

Step 5: Destination Checking – The technique iterates steps 2 to 4 until definite end states are reached, for a specified number of iterations, when ended. The estimation of $\vec{g}_{i,best}$ and $\vec{P}_{i,best}$ give the result.

The fitness values are not considered in PSO algorithms. This is a big computational advantage over other algorithms, when the population is huge. Arithmetic operation of real numbers is used for calculation of velocity and position. The disadvantages as seen in PSO are non-optimal tuning of input features and PSO is one-way information sharing mechanism. In PSO $\vec{g}_{best}$ gives information to others.

3.3.2 WHALE Optimization Algorithm

Lately there has been developing enthusiasm for WOA which was proposed in [10]. This hunt and advancement calculation is a scientific reenactment of the conduct and development of humpback whales as they continued looking for food and arrangements. WOA has motivated by the Bubble-net assaulting system, where the whales begin focusing on fish by making winding formed air pockets around their fish down to 12 meters deep from the surface, and afterward, they swim back up to trap and catch their focused on fish. In light of the general places of whales, in this calculation, the investigation procedure is spoken to by the irregular pursuit of food which can be scientifically interpreted by refreshing the old arrangements as opposed to picking the best ones through haphazardly choosing different arrangements. Notwithstanding this intriguing conduct, WOA is quite recognized from other improvement calculations, since it just needs to modify two parameters. These parameters make it conceivable to change easily between both the abuse and investigation forms

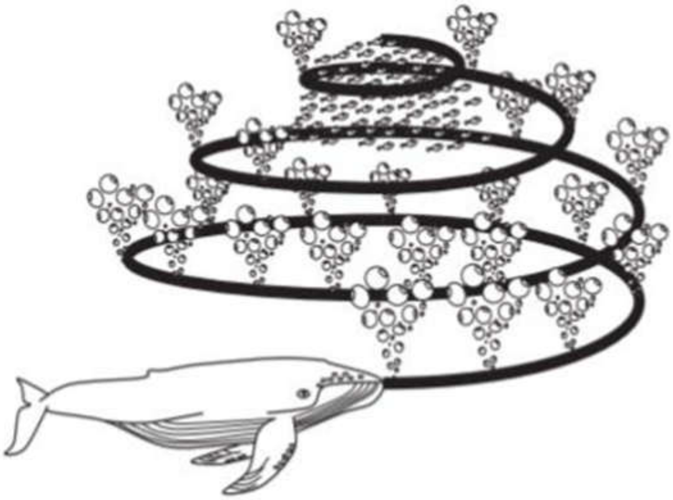


Fig.3. Encircling Attack Prey Searching Methodology for Hump Back Whales

In the following section, we will describe the mathematical model of encircling prey, searching for prey, and spiral bubble-net foraging man oeuvre.

Encircling prey: By the increasing number of iterations from start to a maximum numbers, humpback whales encircle the prey and update their position in the direction of the best search agent. We can mathematically formulate this behavior as:

$$\text{If } (p<0.5 \text{ and mod } (U) <1)$$

Then the position of the candidate position X (t+1) is updated by the following equations

$$D= \text{mod } \{(C.X)-X\ (t)\}$$

$$X\ (t+1) = [X\ (t) – \{U.D\}]$$

Where p =0.1 (constant) X (t+1) is the best position in the current situation. U and D are calculated by the following equations

$$U= \text{mod } \{2.a.r-a\}$$

$$C=2.r$$

Where a is linearly decreases from 2 to 0 and r is the randomly selected vector

Prey Searching: In prey searching mechanism, X is replaced with the random variables X_{random} and mathematical equation are given as follows

$$D= \text{mod } \{(C.X_{random})-X(t)\}$$

$$X(t+1) = [X_{random}\ (t) – \{U.D\}]$$

The encircling the prey and spiral updating of the prey has been done during the exploration phase of whale optimization algorithm. The mathematical expression for updation of new position during the spiral process is given by below equation

$$X(t+1) = D^{l}.e^{bl}.\cos(2\pi l) + X^{*}(t)$$

Where D is the distance between the new position and updated position in new generation, b is the constant which varies from the 0 to 1.

3.4 Machine Learning Algorithms

Machine learning is an integral part of artificial intelligence. This is used to design algorithms based on the data trends and historical relationships between data. Machine learning is used in various fields such as bioinformatics, intrusion detection, information retrieval, game playing, marketing, malware detection, image processing and so on. Machine learning is the most effective method used in the field of classification in order to predict something by developing algorithms. These algorithms allow researchers to produce reliable and valid results and decisions. It also helps to discover some hidden features through historical learning's and trends in data [11]. Feature selection is the most important task of machine learning. Basically machine learning can be grouped into three categories:

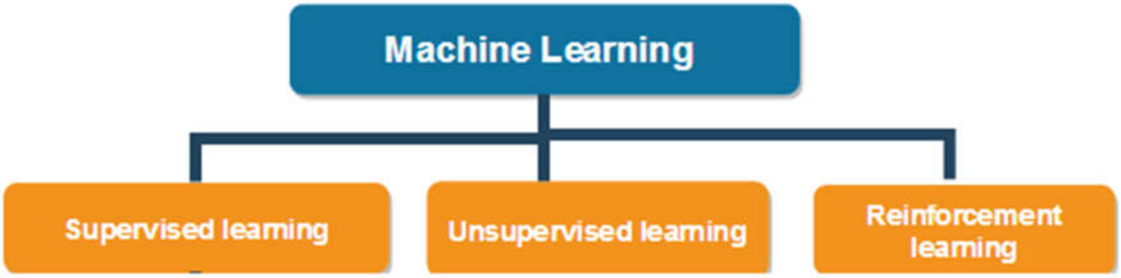


Fig.4. Machine Learning Algorithms

- ***Supervised learning:*** The majority of practical machine learning uses supervised learning. All data is labeled and the algorithms learn to predict the output from the input data. Examples: Support Vector Machines, logistic regression, Naive Bayes, k-nearest neighbor algorithm and Neural Networks (Multilayer perceptron)

- ***Unsupervised learning:*** All data is unlabeled and the algorithms learn to inherent structure from the input data. Examples: K-Means, K-Medoids, Fuzzy C-Means, Hierarchial, Gaussian Mixture, Hidden Markov Model and Neural Networks.

- ***Reinforcement learning:*** Reinforcement learning is all about making decisions sequentially. In simple words, it can define that the out depends on the state of the current input and the next input depends on the output of the previous input. Examples: Decision trees, linear regression, Ensemble methods and Neural Networks.

Choosing the right algorithm is major task in this system. Nowadays, there are lots of supervised and unsupervised machine learning algorithms developed, and each takes a different approach to learning.

3.4.1 ML Algorithms for Classification

ML play a vital role in classification. This section will discuss the some ML classification algorithms such as NB, KNN and RF, LR and demonstrates all classification algorithm's characteristics and working methodology.

a. Naïve Bayes (NB) Algorithm

NB is one of the great ML algorithms, which is utilized for classification process. This method is sustained from Bayes theorem where foundational theory of 'NB' classifier is constructed on the independence theory. ıple spam filtering also other areas of text classifi $= \text{argmax} P\ (c_i)\ P\ (d_r|c_i)$ tures then grades are utilized to approximately compute the probability score of grades of a specified feature subsets. This classifier applied the simple probabilistic classifier, which assist in classifying a data 'd_r', out of classes $c_i \in C$ ($C_{i=1}^{m} = c_1, c_2, \ldots\ldots c_m$). The finest class returns in 'NB' classification is the Maximum Posterior (MAP) class

$$C_{map}$$
$$:1€C$$

Here, the class '$P(c_i)$' can be calculated by dividing the total number of features in class 'c_i' by the entire number of features. P (d_r |c_i) denoted the number of incidence of the feature in data 'd_r' belongs to class 'c_i'. The probability value '$P(c_i| d_r)$' will be calculated for every latent class, but '$P(d_r)$' doesn't change for every class. Accordingly, it can drop the denominator. It chooses the maximum probable classes 'c_{map}' of given data 'd' by computing the posterior probability of every class [12].

b. K-Nearest Neighbor (KNN)

KNN algorithm plays an important role in machine learning system. It belongs to the supervised learning area and have numerous applications in intrusion detection, pattern recognition, and so on. These KNNs are applied in realistic consequences where non-parametric methods are needed.

These techniques do not create any presumptions about data distribution. In the certain dataset, the KNN method classifies the correlatives into clusters which are recognized by a specific characteristic. The most idea for this method is that it's similar output for similar training samples. For the input population nearest value is identified that's ready to assign classes to all or any the samples.

Consider $X_i = \{x_1, x_2, \ldots, x_{iN}\}$ and $X_j = \{x_1, x_2, \ldots, x_{jN}\}$ the sample population, thus to measure the similarity between them and the distance is calculated as given.

$$\mathrm{Dist}(\mathrm{X}_i,\ \mathrm{X}_i) = \sqrt{\sum_{m=1}^{N}\left(\mathrm{x}_{im} - \mathrm{x}_{jm}\right)^2}$$

In the above equation, Euclidean distance is described that evaluates similarity among two pixel points. Hence, the pixels obtain the category to which a number of them commonly resemble [13].

c. Random Forest

Random Forest machine learning algorithm which is flexible and easy to use and also produces great results most of the times. It is widely used because of its simplicity and the fact that it can be both used in classification as well as regression. It functions by building a multitude of decision trees during the training process resulting in the output in the form of classification of individual trees [14].

In the training algorithm of Random Forest, bootstrap aggregation technique is used. Given a training set X= (X_1,X_2,…X_n) with Y= (Y_1,Y_2,…Y_n) bagging B times by selecting a random sample and applies tress to these samples. For b=(1,…..,B). Training Sample Replacement X, Y=(X_b)(Y_b).

Training classification of regression tree of f_b on X_b,Y_b
Predictions (average) from all individual regression trees on x'.

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

Estimate of uncertainty can be made as the standard deviation of the prediction of all individual regression trees on x'.

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B}(f_b(x') - \hat{f}\,)^2}{B-1}}$$

d. Logistic Regression

Logistic regression [7] is widely used to predict a binary response. It is considered a linear method with the loss function in the formula given by the logistic loss by:

$$L\ (w;\, x,\, y) := \log\ (1+\exp\ (-yw^{T}\, x))$$

The algorithm outputs a binary logistic regression model for binary classification problems. If a data point is given, denoted by x, the model can make predictions by application of the following logistic formula:

$$f\ (z) = 1/\ (1+e^{-z}\,)$$

Where z = w^T x. By default, when (w^T x)>0.5, the outcome is positive, else negative. The output of the logistic regression model calculated from the formula f(z), has a probabilistic interpretation (the probability that is positive).The logistic function's formula intakes an input with values ranging from negative infinity to positive infinity, but always outputs values lying in between zero and one and therefore is interpreted as probability measure [15].

4. Data collection Unit

Figure 5 shows the block diagram for the proposed experimentation. The proposed architecture consists of Simulation Phases, Dataset generation, Feature selection and classification models.

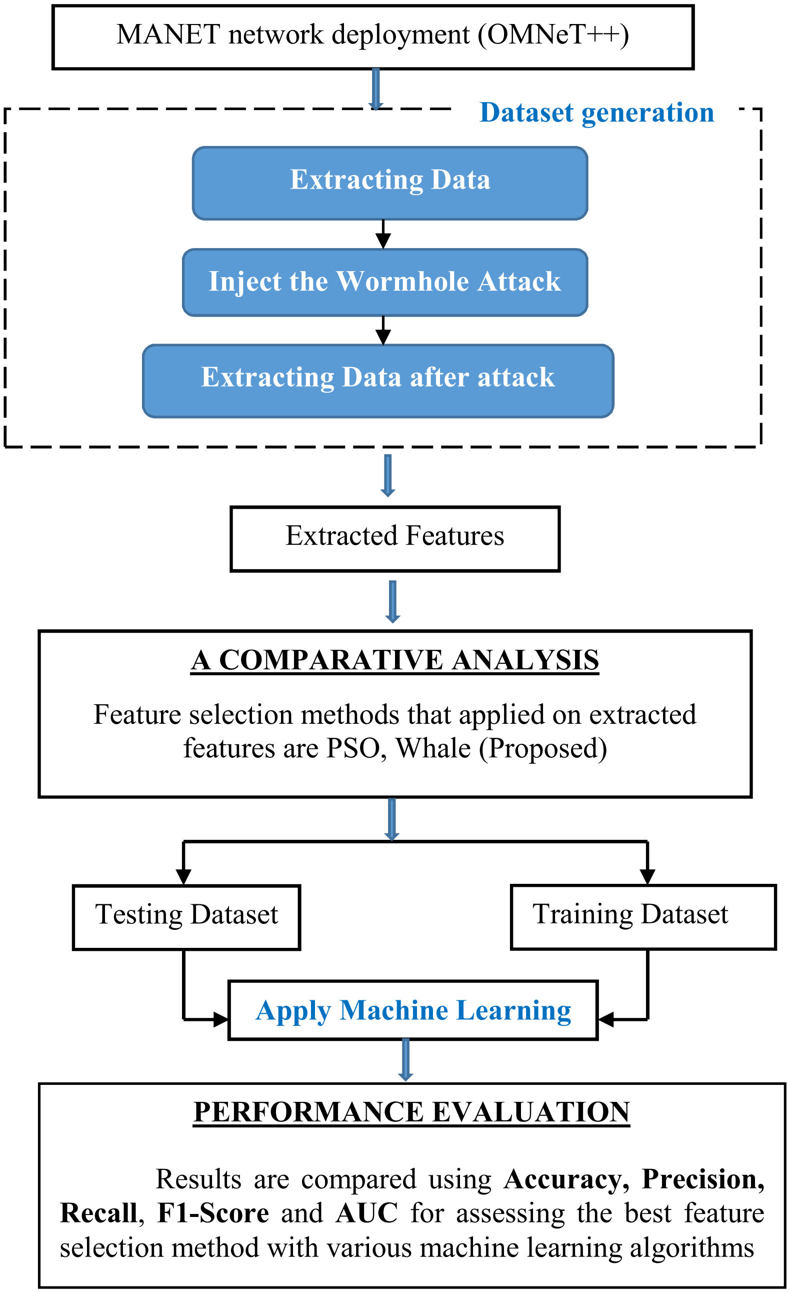

MANET network deployment (OMNeT++)
Dataset generation
Extracting Data
Inject the Wormhole Attack
Extracting Data after attack
Extracted Features
A COMPARATIVE ANALYSIS
Feature selection methods that applied on extracted features are PSO, Whale (Proposed)
Testing Dataset
Training Dataset
Apply Machine Learning
PERFORMANCE EVALUATION
Results are compared using Accuracy, Precision, Recall, F1-Score and AUC for assessing the best feature selection method with various machine learning algorithms

Fig.5. System Architecture

4.1 MANET network deployment (OMNeT++)

Based on the research survey, the most popular open source simulators selected are NS3 and OMNET++ according to that we selected OMNET++ for our experimental analysis in this research. Node distribution scenario is depicted by Fig.2. There are 16 nodes in the network. Simulation parameters are given in Table 1.

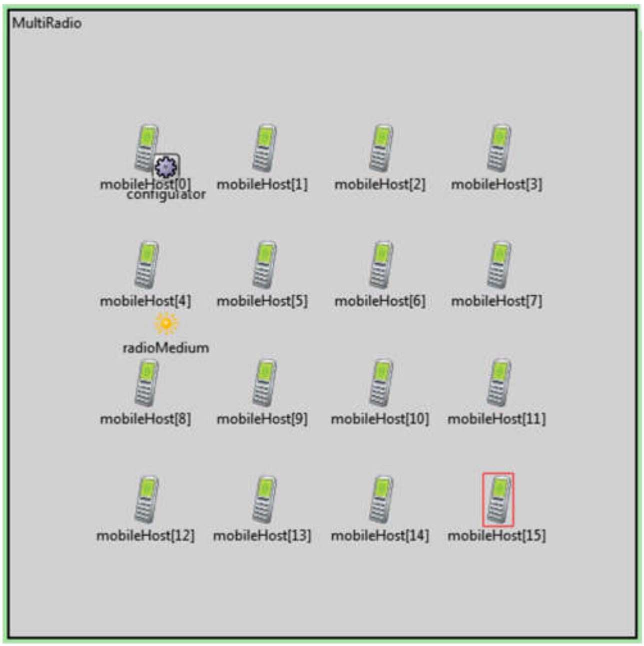


Fig.7. Number of Nodes present in the MANET

- **Data Collection Unit:** Data transmission range is shown in below figure. Wormhole attack is created in among various nodes present in the environment (node 0 to node 15) by creating tunnels.

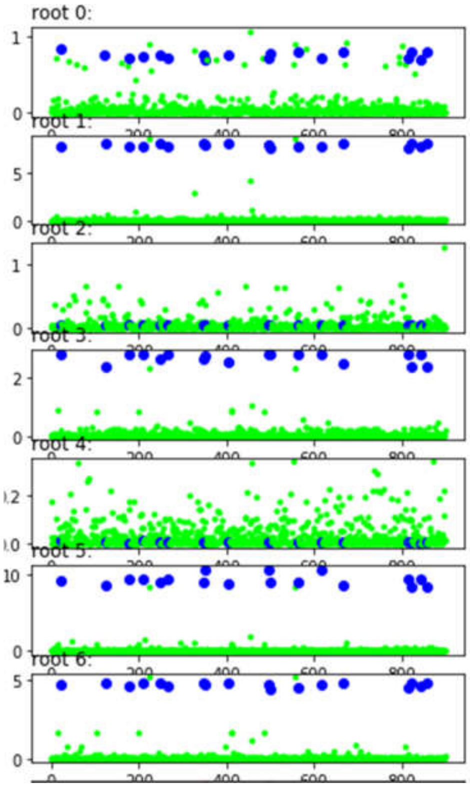
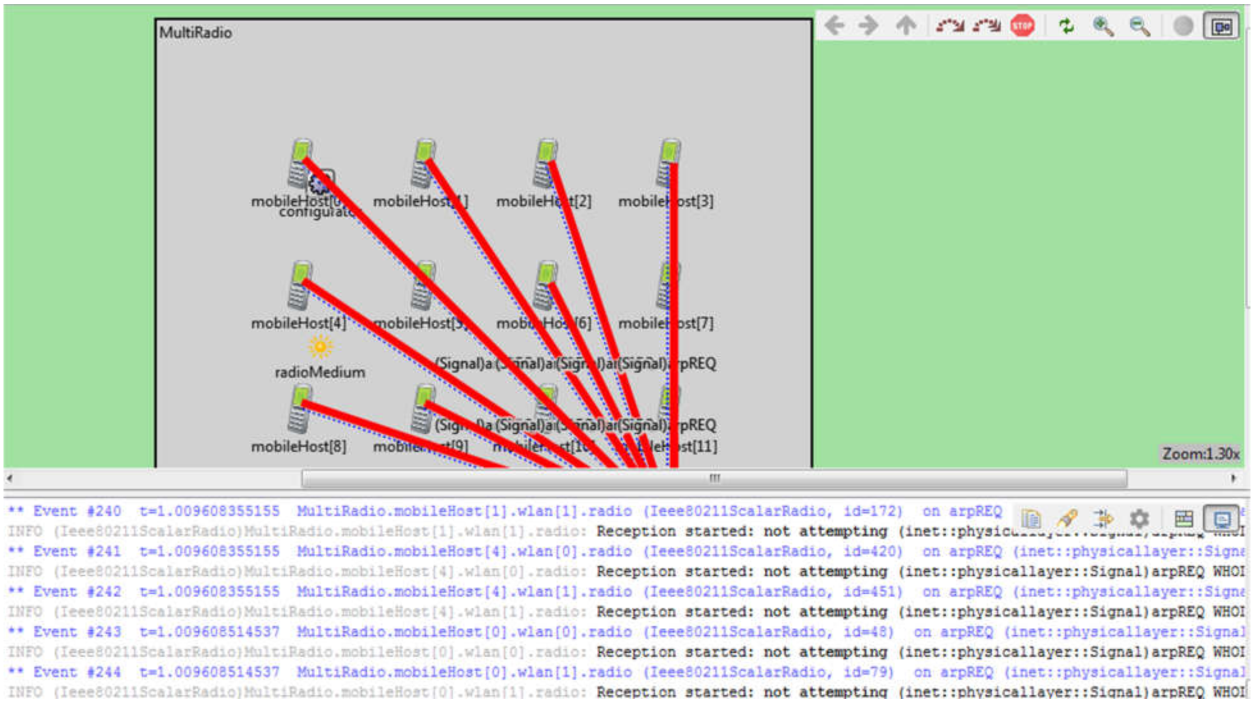


Fig.8. Data transmission range

From the above figure, Normal nodes are depicted by green color and attacked nodes are shown by blue color. Seven root tunnels are created to inject wormhole attack using python tool.



```
** Event #240  t=1.009608355155  MultiRadio.mobileHost[1].wlan[1].radio (Ieee80211ScalarRadio, id=172)  on arpREQ
INFO (Ieee80211ScalarRadio)MultiRadio.mobileHost[1].wlan[1].radio: Reception started: not attempting (inet::physic
** Event #241  t=1.009608355155  MultiRadio.mobileHost[4].wlan[0].radio (Ieee80211ScalarRadio, id=420)  on arpREQ (inet::physicallayer::Signa
INFO (Ieee80211ScalarRadio)MultiRadio.mobileHost[4].wlan[0].radio: Reception started: not attempting (inet::physicallayer::Signal)arpREQ WHOI
** Event #242  t=1.009608355155  MultiRadio.mobileHost[4].wlan[1].radio (Ieee80211ScalarRadio, id=451)  on arpREQ (inet::physicallayer::Signa
INFO (Ieee80211ScalarRadio)MultiRadio.mobileHost[4].wlan[1].radio: Reception started: not attempting (inet::physicallayer::Signal)arpREQ WHOI
** Event #243  t=1.009608514537  MultiRadio.mobileHost[0].wlan[0].radio (Ieee80211ScalarRadio, id=48)  on arpREQ (inet::physicallayer::Signal
INFO (Ieee80211ScalarRadio)MultiRadio.mobileHost[0].wlan[0].radio: Reception started: not attempting (inet::physicallayer::Signal)arpREQ WHOI
** Event #244  t=1.009608514537  MultiRadio.mobileHost[0].wlan[1].radio (Ieee80211ScalarRadio, id=79)  on arpREQ (inet::physicallayer::Signal
INFO (Ieee80211ScalarRadio)MultiRadio.mobileHost[0].wlan[1].radio: Reception started: not attempting (inet::physicallayer::Signal)arpREQ WHOI
```

Fig.9. Dataset generation

Above figure shows that the data generation for all conditions (before and after attack). Features which are generated for classifying/detecting the wormhole attack are listed below,

Table1. Generated features

Event	*Bandwidth*
MultiRadio.mobileHost	*transmission ID*
wlan	*Receiver Id*
Id	*Start Time*
ip	*End Time*
tcpeg	*Preamble Duration*
power	*Data Duration*
Center Frequency	*Header Duration*

To provide an MANET simulation area, we consider a INET framework for autonomous agents performing an exploration task in a defined mission (simulation) area. The list of parameters which are created by INET MANET framework are listed in below Table,

Table2. Simulation parameters in OMNET++/INET MANET environment

Parameters	**Values**
Network Simulator	OMNET++
Medium Access Control (MAC) layer	IEEE802.11g
Number of nodes	16
Simulation area	500 m x 500 m x 250 m
Wireless transmission range	250m
Mobility model	Random waypoint model
Carrier frequency	2.4 GHz
Mobility maximum speed	20 m/Sec
Simulation time per run	300 s
Packet size	512 bytes/packet
Layer Type	Network Layer
Movement speed	50 km/h

5. Results and discussion

5.1. Experimental setup:

As discussed in above section, INET integrated with the OMNET has been used for the data collection under normal and up normal conditions. To identify the wormhole attacks, various feature selection and Machine learning algorithms. Nearly 1000 data were collected and used for the experimenting the proposed algorithm.

5.2 Performance Evaluation

The PSO and whale features which are extracted from the datasets are used for training and testing process. For evaluation, 80% of data were used for training and 20% of data were used for testing. The evaluation is carried out for the different algorithms with the following parameters.

$$\text{Accuracy} = \frac{\text{DR}}{\text{TNI}}\,\text{x}100$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}\quad \text{x}100$$

$$\text{Recall} = \frac{\text{TN}}{\text{TP+FN}}\text{x}100$$

Where TP, TN, FP and FN Represents True Positive, True Negative, False Positive and False Negative values and DR and TNI Represents Number of Detected Results and Total number of Iterations [16]. The performance of the proposed Whale-RF algorithms has been evaluated by different cases which are shown in the below table,

Table3. Performance Evaluation

Description	Algorithm used	Accuracy	Precision	Recall	F1-score	AUC
Machine Learning	LR	94	97	85	89	84
	RF	75	71	85	71	84
	NB	75	71	85	71	84
	KNN	90	88	75	80	75
Optimization with PSO	PSO-LR	95	97	86	91	85
	PSO-RF	81	58	52	49	51
	PSO-NB	81	58	52	49	51
	PSO-KNN	91	93	76	81	76
Optimization with Whale	Whale-LR	95	97	88	92	87
	Whale-RF	98	97	99	95	95
	Whale-NB	84	42	50	46	50
	Whale-KNN	93	85	90	86	88

In the evaluation scenario, wormhole attack is considered for the classification and different comparative analysis are shown in above table.

a. Accuracy analysis

The below figure clearly shows accuracy of the Whale based algorithms has maximum accuracy when compared with the other algorithms.

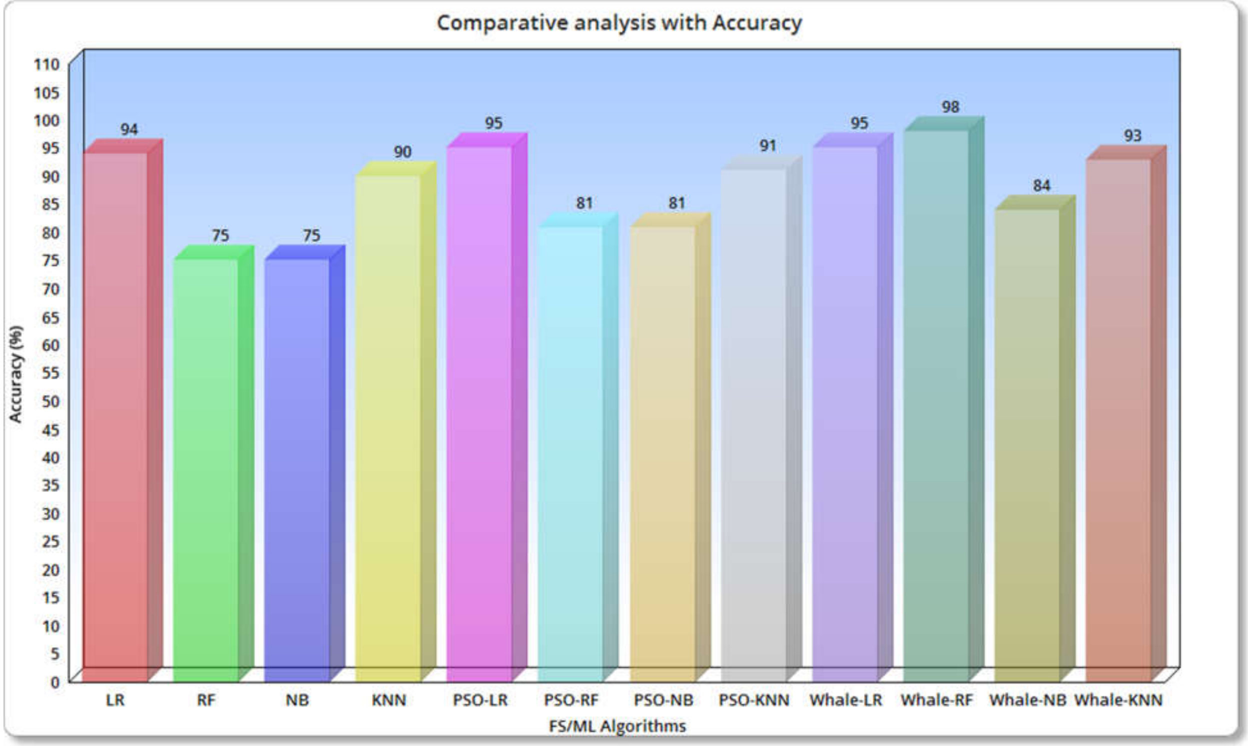


Fig.10. Accuracy analysis for the different algorithms in classifying/predicting the wormhole attack

b. Precision, Recall and F1-Score analysis

Again the precision, recall and F1-Score has been calculated and compared with the other algorithms in which the whale based algorithms outperforms the other algorithms.

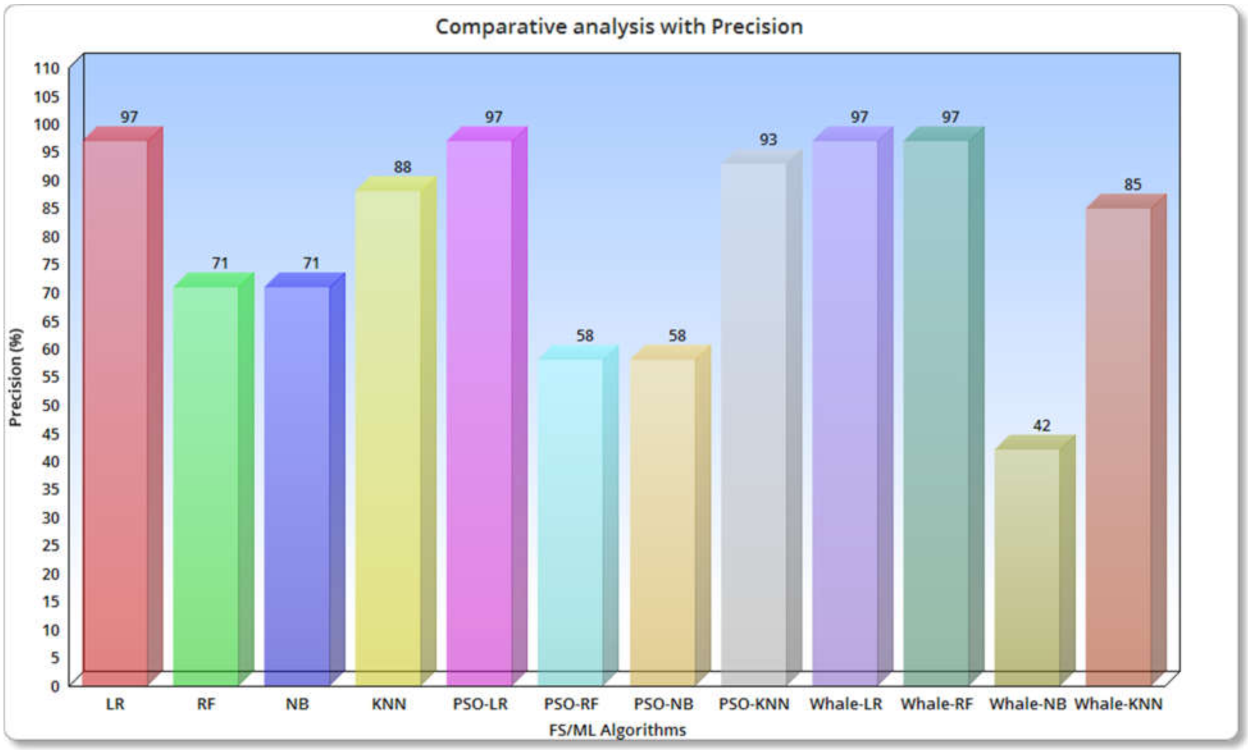


Fig.11. Precision analysis for the different algorithms in classifying/predicting the wormhole attack

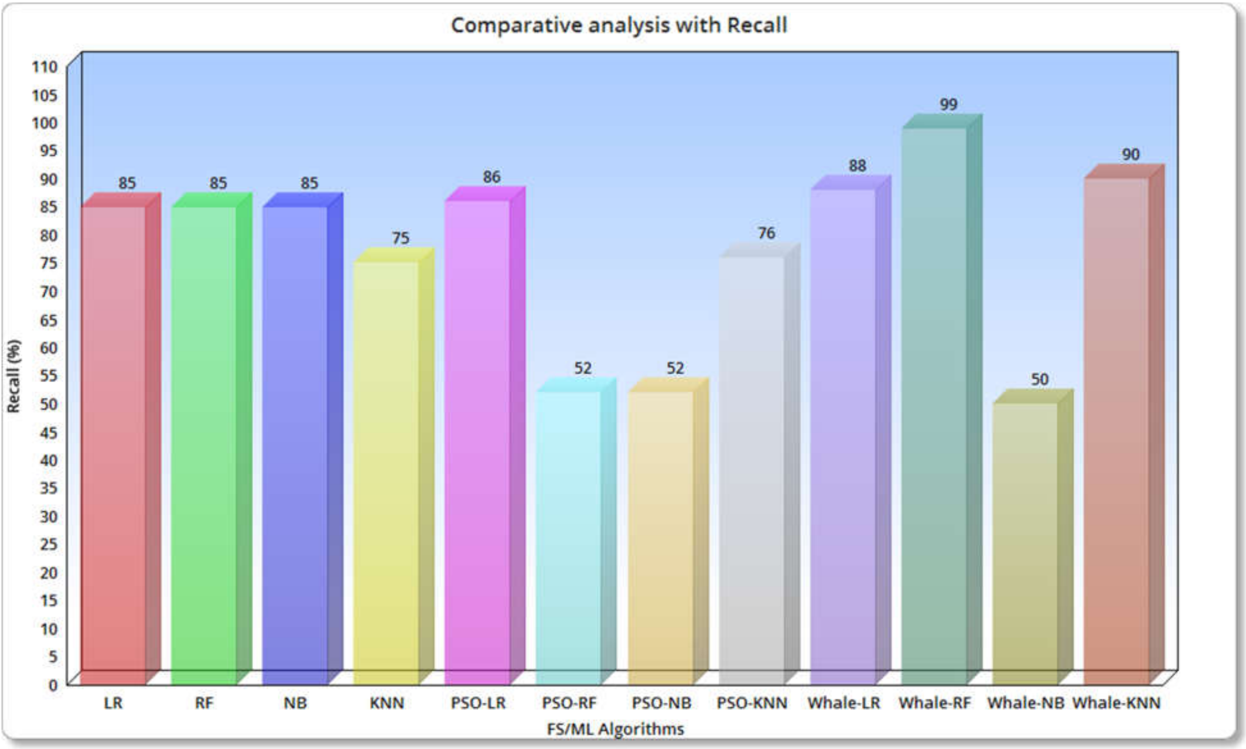


Fig.12. Recall analysis for the different algorithms in classifying/predicting the wormhole attack

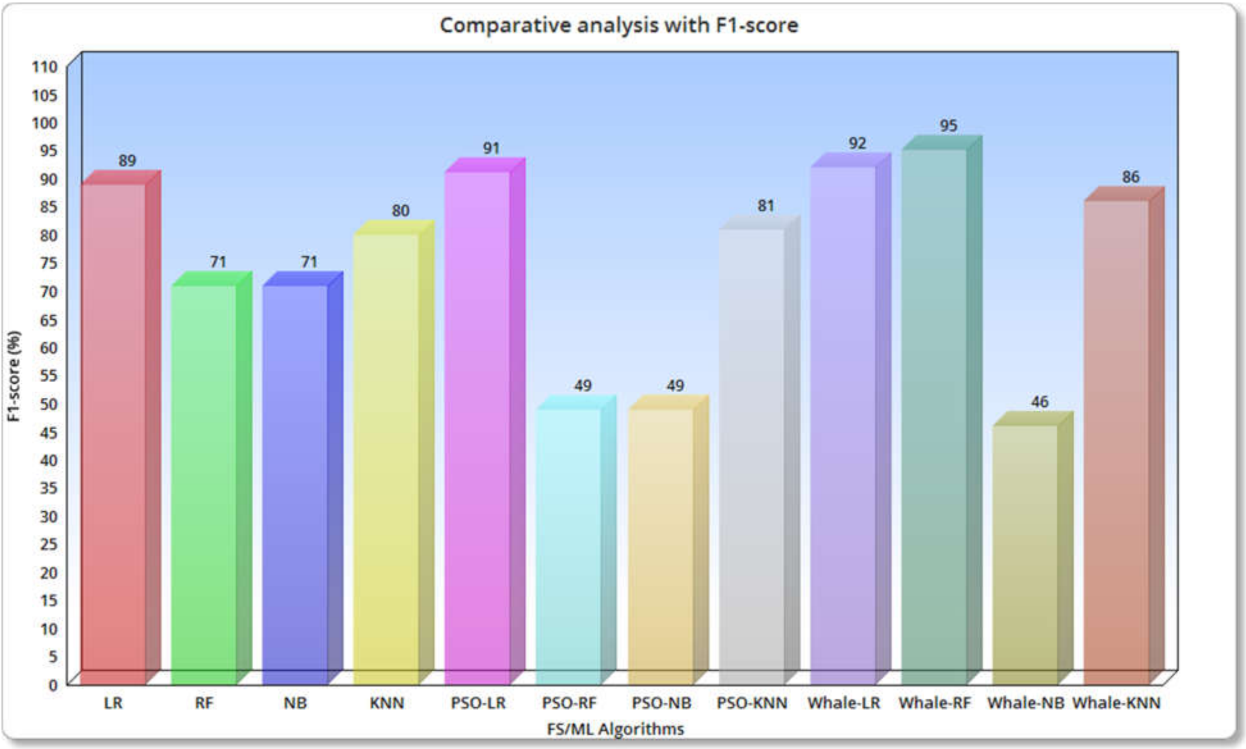


Fig.13. F1-score analysis for the different algorithms in classifying/predicting the wormhole attack

The above clearly shows that whale optimization algorithm gives the beat results when compared with others. The feature selection and ML algorithm provided by the python machine learning toolbox is utilized for assessing the effort of the proposed methodology and calculates the number of wormhole attacks present in the generated dataset. 1000 data were taken for analysis. Among that, 20% (200) has been taken as testing data and the 80% (800) has been taken as training data. With this the accuracy percentage of Whale-RF algorithm has been reached with 98%. According to that, 2% of data has been comes under misclassification scenario.

Actual class	**Predicted class**		
		No (0)	Yes (1)
	No (0)	**167**	**1**
	Yes (1)	**3**	**29**

Fig.14. Confusion Matrix Whale-RF

With the Whale-RF algorithm reached the maximum accuracy when compared with other algorithms. According to that 200 data has been taken as testing data in that 167 cases has been classified under normal nodes. 29 cases has been classified under attacked nodes correctly. The remaining 4 (misclassification) cases may be normal or attacked nodes. Overall, 98% of the predictions are correct remaining 2% could not be predicted accurately (Misclassified).

6. CONCLUSION

Mobile Ad-hoc Network is facing many problems related to the security. A lot of protocols and the algorithmic approaches have been developed for providing the security against the malicious attacks, remove the issues associated to it and to improve the performance of routing protocols. The purpose of this research paper is to explore the ability of feature selection algorithms to detect the wormhole attack in simulated MANET environment. In this research, PSO and Whale methods are developed to select the optimal features for classification process to improve the efficiency of wormhole attack nodes detection in MANET network. The efficiency of classification method is evaluated in terms of Accuracy, F-score, precision, and recall. These experimental results shows that whale algorithm is highly effective and encouraging. The proposed algorithm seems to be more intelligent in detection of the wormhole attacks.

References:

1. Haitham Elwahsh ,Mona Gamal, A. A. Salama, I.M. El-Henawy "A Novel Approach for Classifying MANETs Attacks with a Neutrosophic Intelligent System based on Genetic Algorithm" Hindawi Security and Communication Networks, Volume 2018, Willy.
2. Saurabh Upadhyay, Brijesh Kumar Chaurasia "Detecting and Avoiding Wormhole Attack in MANET Using Statistical Analysis Approach" Institute for Computer Sciences, Social Informatics and Telecommunications Engineering,Springer, 2012.
3. N. Holden and A. A. Freitas, "A hybrid PSO/ACO algorithm for discovering classifcation rules in data mining," Journal of Artifcial Evolution and Applications, vol. 2008, Article ID 316145, 11 pages, 2008.
4. Mohammad Nurul afsar ShaonK. FerensK. Ferens "A Computationally Intelligent Approach to the Detection of Wormhole Attacks in Wireless Sensor Networks" Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 3, 302-320 (2017).
5. Grover, J., Prajapati, N.K., Laxmi, V., Gaur, M.S.: Machine learning approach for multiple misbehavior detection in VANET. In: International Conference on Advances in Computing and Communications, pp. 644–653. Springer (2011).
6. F. Ardjani, K. Sadouni, and M. Benyettou, "Optimization of SVM multiclass by particle swarm (PSO-SVM)," in Proceedings of the 2nd International Workshop on Database Technology and Applications, DBTA2010, 2010.
7. Andrew Sagitta Jauhari , Achmad Imam Kistijantoro "INET FRAMEWORK MODIFICATIONS IN OMNeT++ SIMULATOR FOR MPLS TRAFFIC ENGINEERING" IEEE, 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA).
8. Ram Kishore, Singh Parma Nand "Literature Review of Routing Attacks in MANET" International Conference on Computing, Communication and Automation (ICCCA2016),2016 IEEE.
9. M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis," Knowl.-Based Syst., vol. 125, pp. 116-135, 2017.
10. Hongli Guo,Bin Li,Youmei Zhang,Yu Zhang,Wei Li,Fengjuan Qiao, Xuewen Rong, Shuwang Zhou.Gait Recognition Based on the Feature Extraction of Gabor Filter and Linear Discriminant Analysis and Improved Local Coupled Extreme Learning Machine.Hindawi Mathematical Problems in Engineering Volume 2020.
11. Grover, J., Prajapati, N.K., Laxmi, V., Gaur, M.S.: Machine learning approach for multiple misbehavior detection in VANET. In: International Conference on Advances in Computing and Communications, pp. 644–653. Springer (2011)
12. V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced na¨ıve Bayes model," in Intelligent Data Engineering and Automated Learning –IDEAL 2013, vol. 8206 of Lecture Notes in Computer Science, pp.194–201, Springer, Berlin, Heidelberg, Germany, 2013.
13. Najat Ali, Daniel Neagu, Paul Trundle "Evaluation of K‑Nearest Neighbor Classifier performance for heterogeneous data sets" Springer, SN Applied Sciences, 2019.
14. Angshuman Paul, Dipti Prasad Mukherjee, Prasun Das, Abhinandan Gangopadhyay, Appa Rao Chintha "Improved Random Forest for Classification" IEEE Transactions on Image Processing (Volume: 27 , Issue: 8 , Aug. 2018)
15. Xiaonan Zou, Yong Hu, Zhewen Tian, Kaiyuan Shen "Logistic Regression Model Optimization and Case Analysis" 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)
16. R. Tanuja and A. Umamakeswari, "Efective intrusion detection system design using genetic algorithm for manets," ARPN Journal of Engineering and Applied Sciences, vol. 11, 2016.